A comparison of Chicago neighborhoods in terms of business and social activity against crime rate.

Data Science – Final Capstone Project by Marc Hauptmann

Introduction / Business Problem

- Crime rates in Chicago higher than US average, especially for violent crimes [Ref1]
- 2 related questions to be investigated in this study:
  - Is there a relationship between types of business venue activity in Chicago neighborhoods and the frequency and type of crimes committed?
  - How do neighborhoods in Chicago compare in terms of ‘prosperity’ and ‘safety’ (i.o.w. what are safe, flourishing neighborhoods in Chicago and what distinguishes them from ‘languishing’, ‘crime ridden’ ones)?
- Affected stakeholders:
  - Law enforcement → where to deploy police force?
  - Social and business developers → In which neighborhoods to plan and budget which projects?
  - Business owners and investors → Where to settle which type of businesses?

[Ref1]: http://www.city-data.com/crime/crime-Chicago-Illinois.html, accessed 13-dec 2020.

Data requirements and sources

1. Business venue activity per neighborhood in Chicago , constructed out of:
   1. An overview of Chicago Neighborhoods incl. geospatial definitions of the neighborhood boundaries. [Ref2]
   2. Data on venues incl. types of venues, longitude and latitude, retrieved based on neighborhood location and size. [Ref3]

2. Annual overview of crimes committed in the Chicago area incl. the locations of the crimes committed. [Ref4]

[Ref2]: Neighborhood boundaries in Chicago, developed by the Office of Tourism: https://data.cityofchicago.org/Facilities-Geographic-Boundaries/Boundaries-Neighborhoods/bbvz-uum9 , accessed 13-dec 2020.

[Ref3]: Foursquare website: https://foursquare.com/ , accessed 13-dec 2020.

[Ref4]: https://www.chicago.gov/city/en/dataset/crime.html , accessed 13-dec 2020.

Methodology

- 4 main steps:
  1. Setting up and visualizing the geographical information of the Chicago neighborhoods.
  2. Extracting the relevant crime statistics data, as well as linking to and aggregating with respect to the Chicago Neighborhood geographical information.
  3. Retrieving and analyzing the relevant venue information based on the geographical charac-teristics of the Chicago neighborhoods.
  4. Comparing and analyzing the communalities and differences between the crime and venue characteristics of the Chicago neighborhoods as extracted in steps 2 and 3, respectively.

Methodology con't

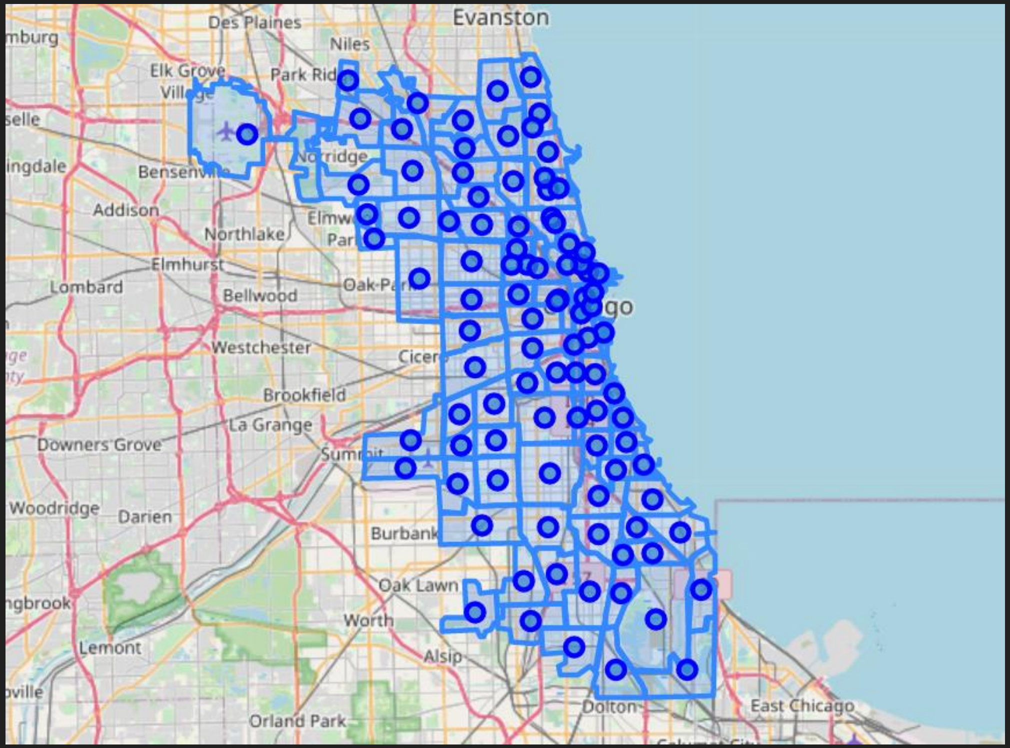
Setting up and visualizing the geographical information of the Chicago neighborhoods

GeoJSON file retrieved via [Ref2]

	pri_neigh	sec_neigh	shape_area	shape_len	geometry
0	Grand Boulevard	BRONZEVILLE	48492503.1554	28196.837157	MULTIPOLYGON (((-87.60671 41.81681, -87.60670 ...
1	Printers Row	PRINTERS ROW	2162137.97139	6864.247156	MULTIPOLYGON (((-87.62761 41.87437, -87.62760 ...
2	United Center	UNITED CENTER	32520512.7053	23101.363745	MULTIPOLYGON (((-87.66707 41.88885, -87.66707 ...
3	Sheffield & DePaul	SHEFFIELD & DEPAUL	10482592.2987	13227.049745	MULTIPOLYGON (((-87.65833 41.92166, -87.65835 ...
4	Humboldt Park	HUMBOLDT PARK	125010425.593	46126.751351	MULTIPOLYGON (((-87.74060 41.88782, -87.74060 ...
5	Garfield Park	GARFIELD PARK	89976069.5947	44460.91922	MULTIPOLYGON (((-87.69540 41.88819, -87.69520 ...
6	North Lawndale	NORTH LAWNDALE	89487422.0244	44959.459663	MULTIPOLYGON (((-87.72024 41.86987, -87.71965 ...
7	Little Village	LITTLE VILLAGE	127998297.819	49904.04003	MULTIPOLYGON (((-87.68740 41.83480, -87.68796 ...
8	Armour Square	ARMOUR SQUARE,CHINATOWN	17141468.6356	24359.189625	MULTIPOLYGON (((-87.62920 41.84713, -87.62919 ...
9	Avalon Park	AVALON PARK,CALUMET HEIGHTS	34852737.7366	27630.822534	MULTIPOLYGON (((-87.58566 41.75150, -87.58475 ...
10	Burnside	CHATHAM,BURNSIDE	16995983.2737	18137.944253	MULTIPOLYGON (((-87.58737 41.72326, -87.58733 ...
11	Hermosa	BELMONT CRAIGIN,HERMOSA	32602059.4055	27179.017438	MULTIPOLYGON (((-87.73146 41.93173, -87.73139 ...
12	Avondale	IRVING PARK,AVONDALE	55290595.482	34261.933404	MULTIPOLYGON (((-87.68799 41.93610, -87.68798 ...
13	Logan Square	LOGAN SQUARE	71256806.5518	36361.508518	MULTIPOLYGON (((-87.68732 41.91399, -87.68751 ...
14	Calumet Heights	AVALON PARK,CALUMET HEIGHTS	48826241.5643	32925.365871	MULTIPOLYGON (((-87.54568 41.72282, -87.54600 ...
15	East Side	SOUTHEAST SIDE	83241728.0493	52274.192849	MULTIPOLYGON (((-87.52462 41.69180, -87.52501 ...
16	West Pullman	WEST PULLMAN	99365198.0822	50023.843001	MULTIPOLYGON (((-87.61828 41.65911, -87.61829 ...
17	Garfield Ridge	MIDWAY AIRPORT	117890778.429	60080.44797	MULTIPOLYGON (((-87.73856 41.81871, -87.73853 ...
18	New City	BACK OF THE YARDS	134636963.254	48133.595961	MULTIPOLYGON (((-87.63546 41.79448, -87.63599 ...
19	Englewood	ENGLEWOOD	173600015.009	56144.046048	MULTIPOLYGON (((-87.62854 41.79414, -87.62853 ...

1. Visualize Neighborhood boundaries in Folium.

2. Determine centroids with geopandas and visualize as markers in Folium



Methodology con’t

Extracting the relevant crime statistics data, as well as linking to and aggregating with respect to the Chicago Neighborhood geographical information.

Crime data from [Ref4]

1. Map to neighborhoods using geopandas

Dataframe with description of crime and mapped neighborhood

	PRIMARY DESCRIPTION	SECONDARY DESCRIPTION	LATITUDE	LONGITUDE	pri_neigh
	NARCOTICS	POSSESS - CRACK	41.819174	-87.628125	Grand Boulevard
	DECEPTIVE PRACTICE	BOGUS CHECK	41.810546	-87.606588	Grand Boulevard
	ROBBERY	STRONG ARM - NO WEAPON	41.805121	-87.621104	Grand Boulevard
	CRIMINAL DAMAGE	TO VEHICLE	41.803121	-87.609460	Grand Boulevard
	BATTERY	DOMESTIC BATTERY SIMPLE	41.808553	-87.622816	Grand Boulevard

2. Aggregate

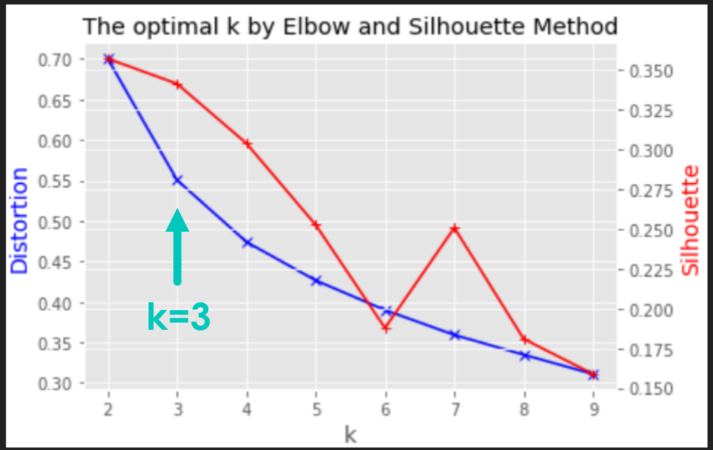
Crime rate per neighborhood

	Neighborhood	CrimeCount
0	Austin	11997
1	Englewood	10930
2	Garfield Park	8616
3	South Shore	7656
4	Humboldt Park	7540
5	North Lawndale	7032
6	Auburn Gresham	6249
7	Roseland	5917
8	Grand Crossing	5754
9	Chatham	5180

3. Determine frequency per crime type per neighborhood

	Neighborhood	ARSON: AGGRAVATED	ARSON: ATTEMPT ARSON	ARSON: BY EXPLOSIVE	ARSON: BY FIRE	ARSON: POSSESSION - EXPLOSIVE / INCENDIARY DEVICE	ASSAULT: AGG PO HANDS NO/MIN INJURY	ASSAULT: AGG PRO.EMP: HANDGUN	ASSAULT: AGG PRO.EMP: OTHER DANG WEAPON	ASSAULT: AGG PRO.EMP: OTHER FIREARM	...	WEAPONS VIOLATION: UNLAWFUL SALE - HANDGUN	WEAP(VIOLATI UNLAW S. HAND(
0	Albany Park	0.000000	0.000513	0.0	0.001026	0.0	0.000000	0.0	0.0	0.0	...	0.000000	
1	Andersonville	0.000000	0.000000	0.0	0.000000	0.0	0.000000	0.0	0.0	0.0	...	0.000000	
2	Archer Heights	0.000000	0.000000	0.0	0.004418	0.0	0.000000	0.0	0.0	0.0	...	0.000000	
3	Armour Square	0.000000	0.000000	0.0	0.000000	0.0	0.000000	0.0	0.0	0.0	...	0.000000	
4	Ashburn	0.000000	0.000000	0.0	0.003808	0.0	0.000544	0.0	0.0	0.0	...	0.000000	

4. kMeans clustering with “optimal” number of groups k



Methodology con’t

Retrieving and analyzing the relevant venue information based on the geographical characteristics of the Chicago neighborhoods.

Venue information extracted from [Ref3] and mapped to each neighborhood

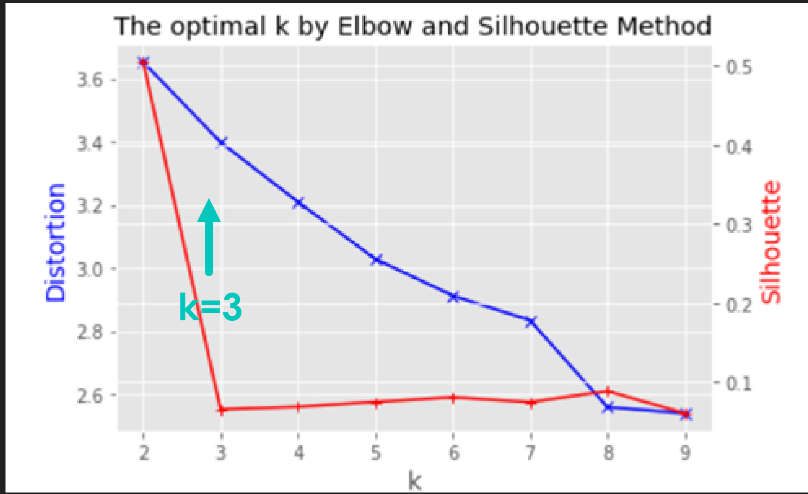
	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category	geometry
0	Grand Boulevard	41.812949	-87.617860	Chicago Blues District	41.810071	-87.614105	Jazz Club	POINT (-87.61411 41.81007)
1	Grand Boulevard	41.812949	-87.617860	Blues Brothers Mural / Shelly's Loan & Jewelry...	41.809391	-87.619517	Plaza	POINT (-87.61952 41.80939)
2	Grand Boulevard	41.812949	-87.617860	Peach's Restaurant	41.809481	-87.617009	Breakfast Spot	POINT (-87.61701 41.80948)
3	Grand Boulevard	41.812949	-87.617860	Ain't She Sweet Cafe	41.816817	-87.613004	Coffee Shop	POINT (-87.61300 41.81682)
4	Grand Boulevard	41.812949	-87.617860	Parkway Ballroom	41.813142	-87.616064	Food	POINT (-87.61606 41.81314)

1. Determine frequency per venue type per neighborhood

	Neighborhood	Yoga Studio	ATM	Accessories Store	Adult Boutique	Afghan Restaurant	African Restaurant	Airport	Airport Food Court	Airport Lounge	...	Warehouse	Warehouse Store	Waterfront	Wei L Cei
0	Albany Park	0.000000	0.0	0.000000	0.0	0.000000	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.000
1	Andersonville	0.000000	0.0	0.000000	0.0	0.000000	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.000
2	Archer Heights	0.000000	0.0	0.000000	0.0	0.000000	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.000
3	Armour Square	0.000000	0.0	0.000000	0.0	0.000000	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.000
4	Ashburn	0.000000	0.0	0.000000	0.0	0.000000	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.020

2. kMeans clustering with “optimal” number of groups k

Use same k as for crime-based clustering for bettercomparison



Methodology con’t

Comparing and analyzing the communalities and differences between the crime and venue characteristics of the Chicago neighborhoods.

Global KPIs for clustering comparison:

- Adjusted Rand Index [Ref5]
- Jaccard similarity score [Ref6]:
  - Requires remapping of clusters
  - Weighted average of per cluster score
  - Total score based on total number of true and false positives and negatives

Confusion matrix [Ref7]:

Venue-based clusters

Crime-based clusters

	Cluster 0	Cluster 1	Cluster 2	Total
Cluster 0	#Mat-ches	#Mat-ches	#Mat-ches	**RowSum**
Cluster 1	#Mat-ches	#Mat-ches	#Mat-ches	**RowSum**
Cluster 2	#Mat-ches	#Mat-ches	#Mat-ches	**RowSum**
Total	**ColSum**	**ColSum**	**ColSum**	

[Ref5]: "Objective criteria for the evaluation of clustering methods" by William M. Rand, Journal of the American Statistical Association 66 (336): 846–850 (1971).

[Ref6]: https://scikit-learn.org/stable/modules/generated/sklearn.metrics.jaccard_score.html , accessed 3-jan 2021.

[Ref7]: “Selecting and interpreting measures of thematic classification accuracy" by Stephen V. Stehman, Remote Sensing of Environment. 62 (1): 77–89 (1997).

Results

1. Crime characteristics of Chicago neighborhoods

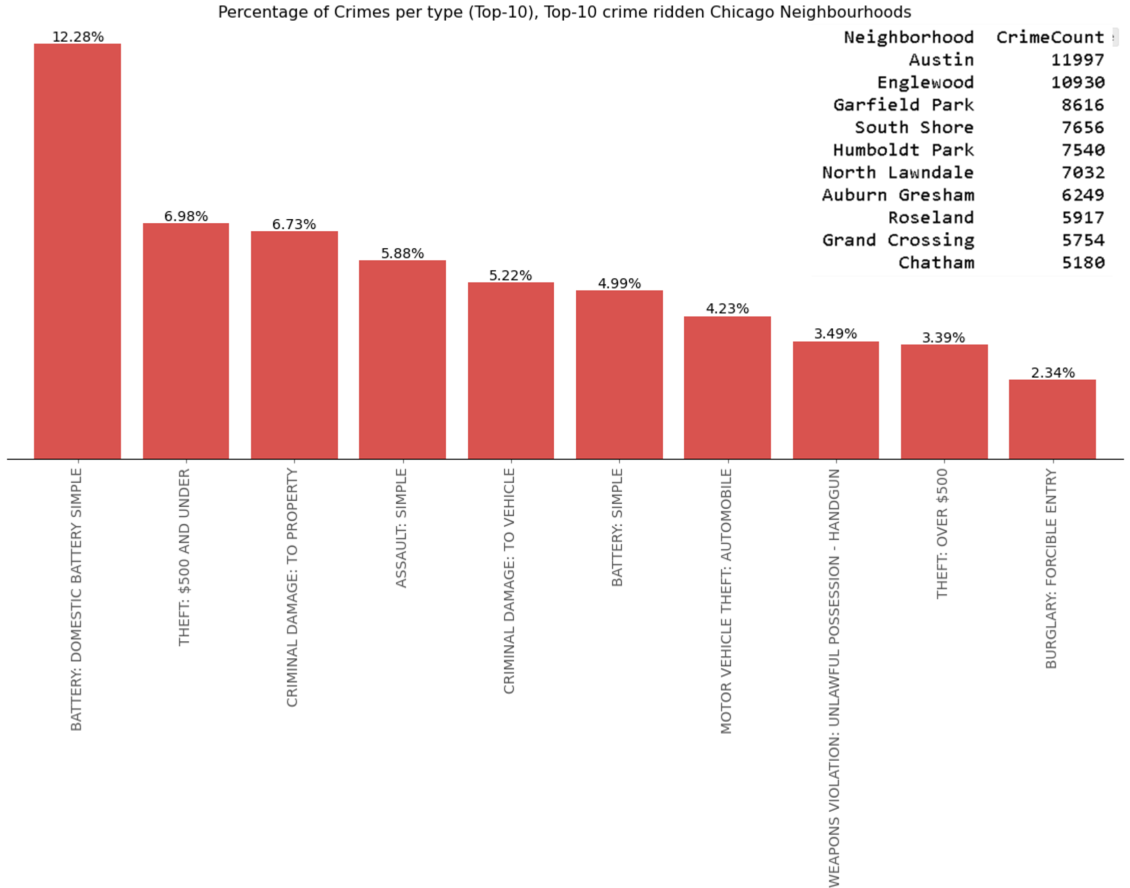
2. Characteristic venues of Chicago neighborhoods

3. Relationship between crime and venue characteristics of Chicago neighborhoods

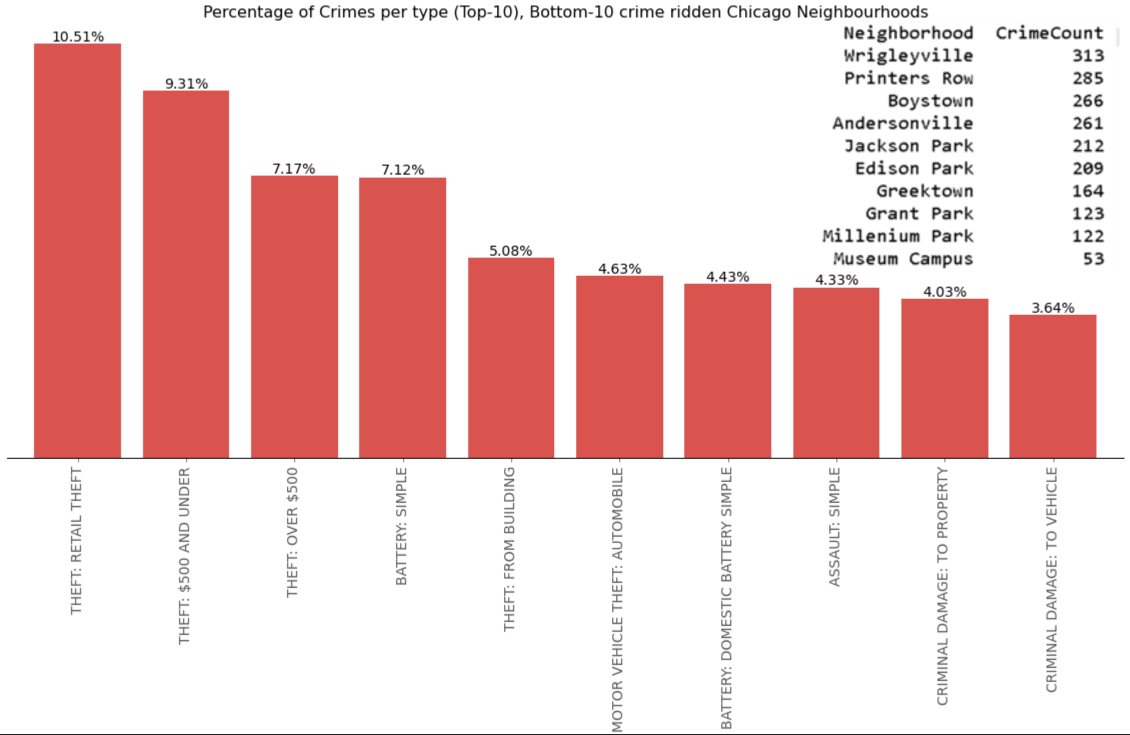
Results con't

Crime characteristics of Chicago neighborhoods

Top crime-ridden neighborhoods characterized by violent crimes



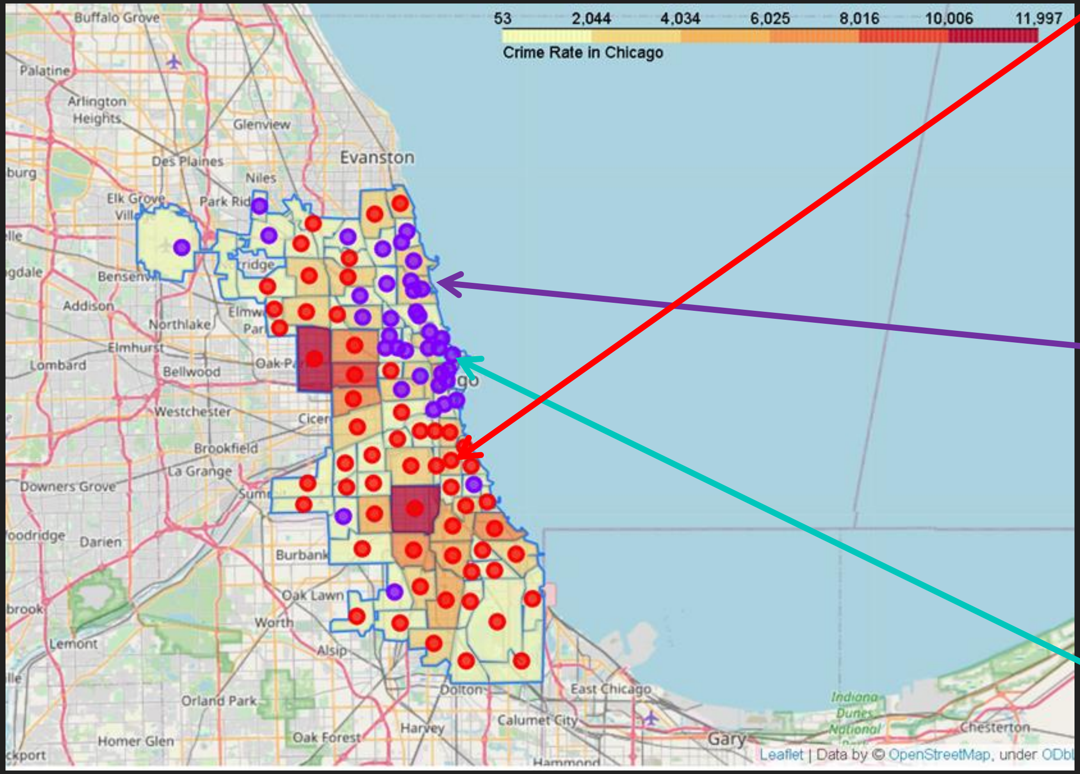
Least crime-ridden neighborhoods characterized by various types of theft



Results con't

Crime characteristics of Chicago neighborhoods

Geographic distribution of different types of crimes and link to crime rate



Violent crimes

pri_neigh	Silhouette	1st Most Common Crime	2nd Most Common Crime	3rd Most Common Crime	4th Most Common Crime	5th Most Common Crime	6th Most Common Crime	7th Most Common Crime	8th Most Common Crime	9th Most Common Crime	10th Most Common Crime
Chicago Lawn	0.609052	BATTERY: DOMESTIC BATTERY SIMPLE	THEFT: $500 AND UNDER	CRIMINAL DAMAGE: TO PROPERTY	ASSAULT: SIMPLE	CRIMINAL DAMAGE: TO VEHICLE	BATTERY: SIMPLE	THEFT: OVER $500	MOTOR VEHICLE THEFT: AUTOMOBILE	BURGLARY: FORCIBLE ENTRY	WEAPONS VIOLATION: RECKLESS FIREARM DISCHARGE
South Shore	0.597854	BATTERY: DOMESTIC BATTERY SIMPLE	CRIMINAL DAMAGE: TO PROPERTY	THEFT: $500 AND UNDER	ASSAULT: SIMPLE	CRIMINAL DAMAGE: TO VEHICLE	BATTERY: SIMPLE	MOTOR VEHICLE THEFT: AUTOMOBILE	THEFT: OVER $500	BURGLARY: FORCIBLE ENTRY	OTHER OFFENSE: TELEPHONE THREAT
Grand Crossing	0.596157	BATTERY: DOMESTIC BATTERY SIMPLE	CRIMINAL DAMAGE: TO PROPERTY	THEFT: $500 AND UNDER	ASSAULT: SIMPLE	MOTOR VEHICLE THEFT: AUTOMOBILE	CRIMINAL DAMAGE: TO VEHICLE	BATTERY: SIMPLE	WEAPONS VIOLATION: UNLAWFUL POSSESSION - HANDGUN	THEFT: OVER $500	BURGLARY: FORCIBLE ENTRY

Petty and heavier theft

pri_neigh	Silhouette	1st Most Common Crime	2nd Most Common Crime	3rd Most Common Crime	4th Most Common Crime	5th Most Common Crime	6th Most Common Crime	7th Most Common Crime	8th Most Common Crime	9th Most Common Crime	10th Most Common Crime
Lake View	0.360684	THEFT: $500 AND UNDER	THEFT: RETAIL THEFT	THEFT: OVER $500	THEFT: FROM BUILDING	BATTERY: SIMPLE	MOTOR VEHICLE THEFT: AUTOMOBILE	ASSAULT: SIMPLE	CRIMINAL DAMAGE: TO PROPERTY	DECEPTIVE PRACTICE: FINANCIAL IDENTITY THEFT $...	CRIMINAL DAMAGE: TO VEHICLE
Lincoln Park	0.349276	THEFT: $500 AND UNDER	THEFT: OVER $500	THEFT: RETAIL THEFT	THEFT: FROM BUILDING	BATTERY: SIMPLE	CRIMINAL DAMAGE: TO VEHICLE	CRIMINAL DAMAGE: TO PROPERTY	ASSAULT: SIMPLE	BATTERY: DOMESTIC BATTERY SIMPLE	BURGLARY: FORCIBLE ENTRY
Wicker Park	0.330413	THEFT: OVER $500	THEFT: RETAIL THEFT	THEFT: $500 AND UNDER	BURGLARY: FORCIBLE ENTRY	THEFT: FROM BUILDING	BATTERY: SIMPLE	ASSAULT: SIMPLE	CRIMINAL DAMAGE: TO PROPERTY	MOTOR VEHICLE THEFT: AUTOMOBILE	BATTERY: DOMESTIC BATTERY SIMPLE

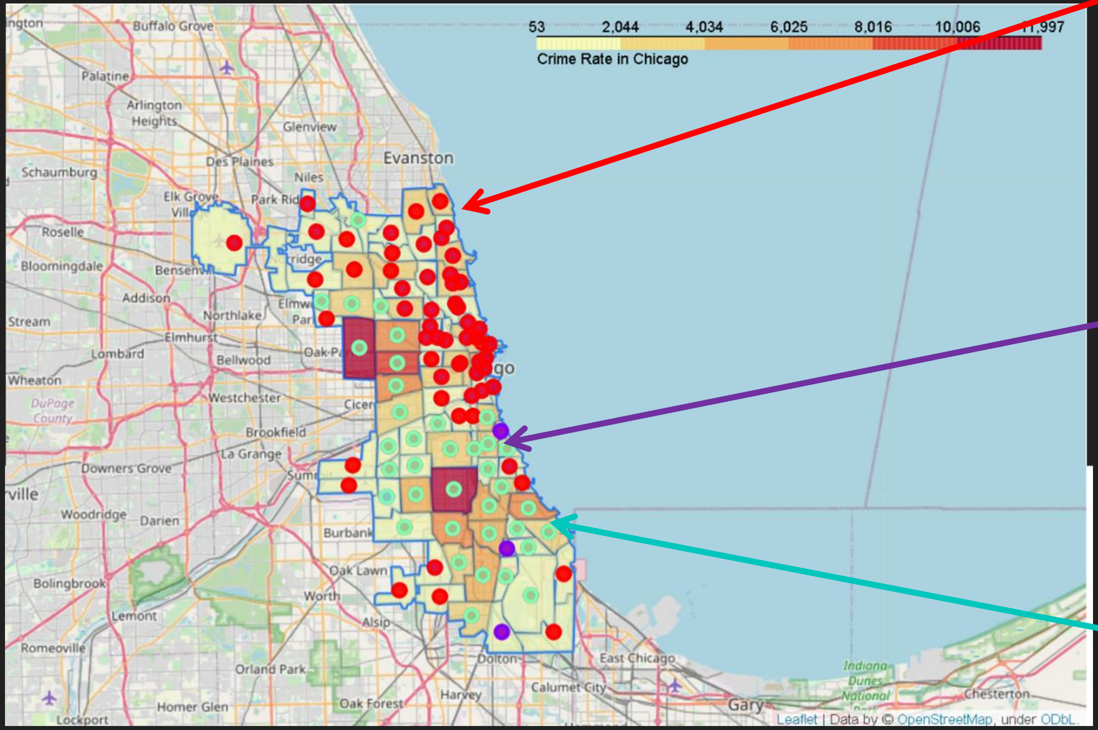
Retail theft

pri_neigh	Silhouette	1st Most Common Crime	2nd Most Common Crime	3rd Most Common Crime	4th Most Common Crime	5th Most Common Crime	6th Most Common Crime	7th Most Common Crime	8th Most Common Crime	9th Most Common Crime	10th Most Common Crime
Magnificent Mile	0.636912	THEFT: RETAIL THEFT	BURGLARY: FORCIBLE ENTRY	THEFT: FROM BUILDING	DECEPTIVE PRACTICE: CREDIT CARD FRAUD	THEFT: OVER $500	THEFT: $500 AND UNDER	THEFT: POCKET-PICKING	BATTERY: SIMPLE	CRIMINAL DAMAGE: TO PROPERTY	ASSAULT: SIMPLE
Greektown	0.563680	THEFT: RETAIL THEFT	THEFT: $500 AND UNDER	BATTERY: SIMPLE	THEFT: OVER $500	ASSAULT: SIMPLE	THEFT: FROM BUILDING	CRIMINAL DAMAGE: TO PROPERTY	MOTOR VEHICLE THEFT: AUTOMOBILE	BURGLARY: FORCIBLE ENTRY	BATTERY: DOMESTIC BATTERY SIMPLE

Results con't

Characteristic venues of Chicago neighborhoods

Geographic distribution of different types of venues and link to crime rate



pri_neigh	Silhouette	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
West Loop	0.112451	Coffee Shop	Italian Restaurant	Café	Dance Studio	Deli / Bodega	Grocery Store	Pizza Place	Bar	New American Restaurant	Gym / Fitness Center
Wicker Park	0.109663	Pizza Place	Bakery	Italian Restaurant	Bar	Coffee Shop	Sushi Restaurant	Bookstore	Nail Salon	Salon / Barbershop	Gourmet Shop
Bucktown	0.104620	Bar	Coffee Shop	Hot Dog Joint	French Restaurant	Gym	Korean Restaurant	Grocery Store	Mexican Restaurant	Park	Pizza Place

Bars, Restaurants,

pri_neigh	Silhouette	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
Burnside	0.175612	Park	Intersection	Train Station	North Indian Restaurant	Pakistani Restaurant	Outlet Store	Outdoors & Recreation	Outdoor Sculpture	Other Repair Shop	Other Great Outdoors
Riverdale	-0.041991	Park	Harbor / Marina	Residential Building (Apartment / Condo)	Garden	Grocery Store	North Indian Restaurant	Outlet Store	Outdoors & Recreation	Outdoor Sculpture	Other Repair Shop
Oakland	-0.117170	Beach	Park	Track	Hotel	Vineyard	Juice Bar	Dog Run	Athletics & Sports	Climbing Gym	Public Art

Outdoor

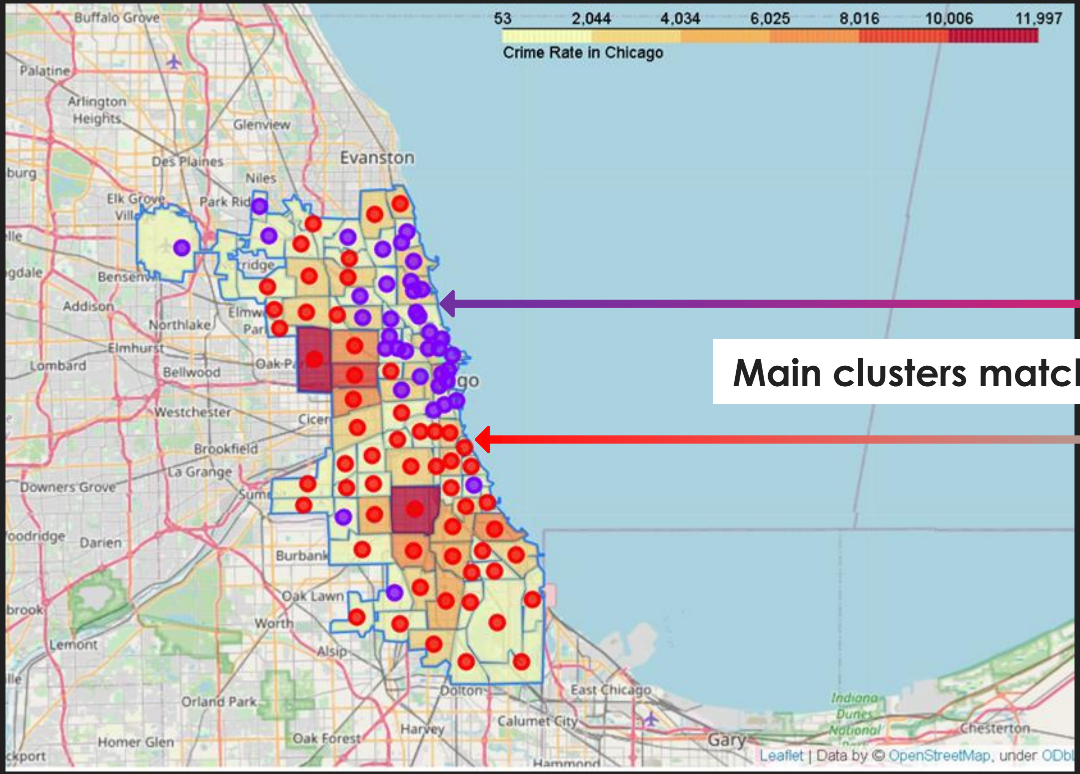
pri_neigh	Silhouette	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
Austin	0.238561	Fast Food Restaurant	Discount Store	Park	Sandwich Place	Fried Chicken Joint	Grocery Store	Donut Shop	Pharmacy	Mexican Restaurant	Clothing Store
Roseland	0.217754	Fast Food Restaurant	Fried Chicken Joint	Sandwich Place	Donut Shop	Discount Store	Liquor Store	Park	Shop & Service	Grocery Store	Food
Englewood	0.213879	Fast Food Restaurant	Discount Store	Sandwich Place	Donut Shop	Gas Station	Grocery Store	Clothing Store	Fried Chicken Joint	Bank	Shoe Store

Fast food, Discount stores

Results con't

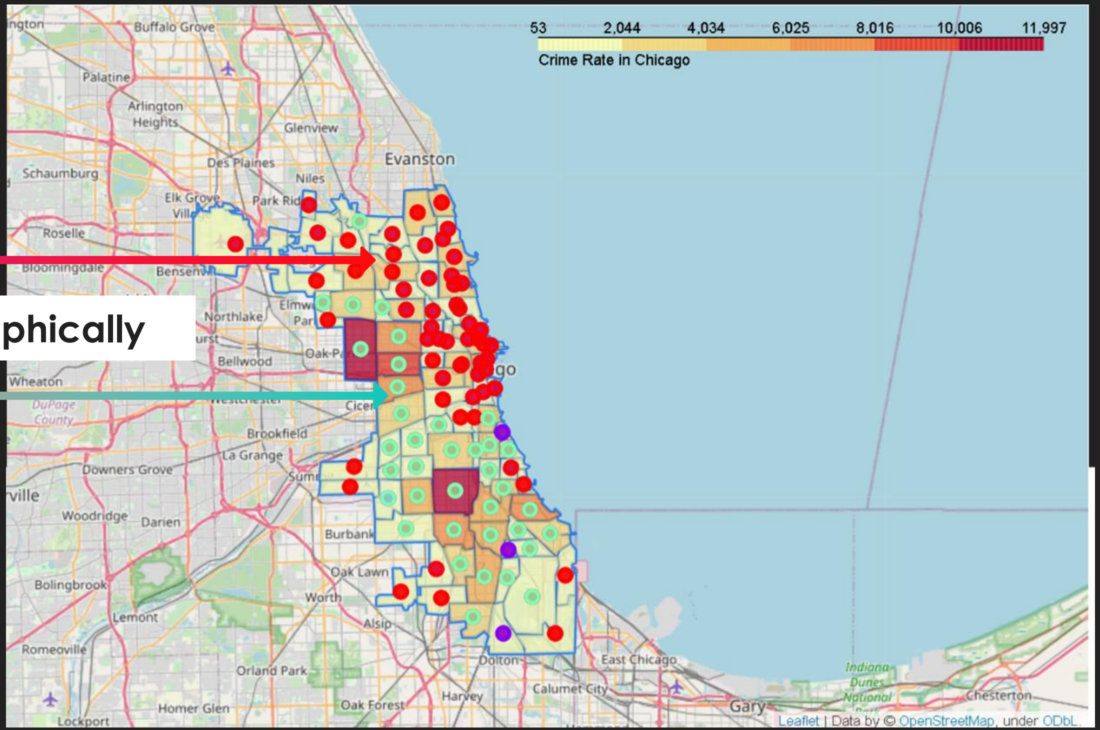
Relationship between crime and venue characteristics of Chicago neighborhoods.

Geographic distribution of different types of crimes and link to crime rate



Main clusters match geographically

Geographic distribution of different types of venues and link to crime rate



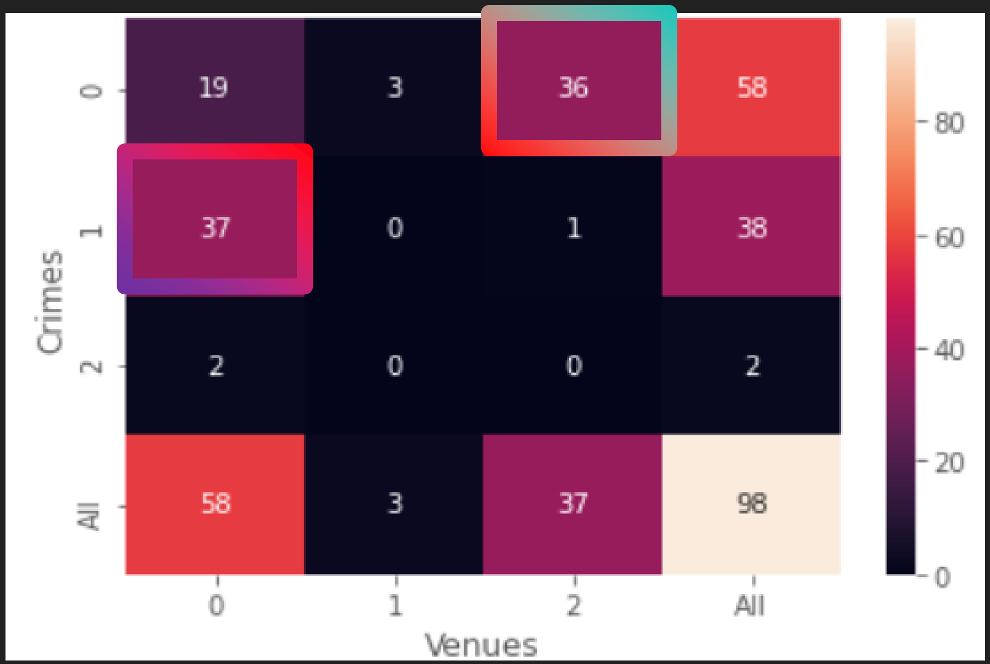
Results con't

Relationship between crime and venue characteristics of Chicago neighborhoods.

Global KPIs for clustering comparison:

- Adjusted Rand Index: **0.27**
  - Affected by cluster size
- Jaccard similarity score:
  - Weighted average of per cluster score **0.60**
  - Total score based on total number of true and false positives and negatives **0.59**

Confusion matrix:



Main clusters match in terms of neighborhoods assigned

Discussion

- Prosperous neighborhoods (North-East of Chicago) exhibit vast amount of high-profile business venues
  - Crime rate generally lower in these neighborhoods, with the committed crimes being less severe.

- Less prosperous neighborhoods ([South]-East of Chicago) dominated by less glamorous venues.
  - Crime rate is the highest in Chicago, with the majority of crimes of more violent nature.

Discussion con't

- The likely underlying rootcause can be found in the socio-demographic characteristics of these neighborhoods:
  - The majority of the population in the crime-ridden neighborhoods lives in precarious circumstances.
    - → less attractive for more high-profile business ventures due to a lack of matching "clientele". [Ref8-10]
    - → lack of employment and perspective may cause high amount of (domestic) violence.
  - More prosperous neighborhoods are either mainly offering high profile housing like lofts [Ref11] or are in-fact non-residential business and/or entertainment areas [Ref12-14].
    - → These may in general attract more wealthy "clientele"

[Ref8]: https://en.wikipedia.org/wiki/Austin,_Chicago , accessed 3-jan 2021.
[Ref9]: https://en.wikipedia.org/wiki/Englewood,_Chicago , accessed 3-jan 2021.
[Ref10]: https://en.wikipedia.org/wiki/West_Garfield_Park,_Chicago & https://en.wikipedia.org/wiki/East_Garfield_Park,_Chicago , accessed 3-jan 2021.
[Ref11]: https://en.wikipedia.org/wiki/Printer%27s_Row,_Chicago , accessed 3-jan 2021.
[Ref12]: https://en.wikipedia.org/wiki/Millennium_Park , accessed 3-jan 2021.
[Ref13]: https://en.wikipedia.org/wiki/Museum_Campus , accessed 3-jan 2021.
[Ref14]: https://en.wikipedia.org/wiki/Magnificent_Mile , accessed 3-jan 2021.

Discussion con't

- In general little diversity in terms of venues → Recommendation to develop neighborhoods appropriately in order to attract appropriate businesses, increase quality of living, while avoiding repelling current population.

Conclusion and outlook

- Characteristics that separate poor, crime-ridden neighborhoods from flourishing, ‘safe’ neighborhoods identified.
- Some general suggestions provided on how (not) to try to improve the situation in the poorer, more crime-ridden areas in a sustainable, long-term way.
- Future investigations to yield more in-depth insights and specific recommendations should include more data sources and aspects:
  - E.g.: population demographics, educational situation, other types of business activity (e.g. industrial activity), etc.
  - Use more elaborate machine learning techniques to maximize feature co-variance such as PCA. [Ref15]

[Ref15]: “Analysis of a complex of statistical variables into principal components” by Harold Hotelling, Journal of Educational Psychology, 24, 417–441, and 498–520, (1933).